DOCUMENT RESUME

ED 392 826 TM 024 492

AUTHOR Rose, Howard
TITLE Assessing Learning in VR: Towards Developing a Paradigm. Virtual Reality Roving Vehicles (VRRV) Project.
SPONS AGENCY U.S. West Foundation.
PUB DATE 16 Feb 95
NOTE 47p.
PUB TYPE Reports - Descriptive (141) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes; *Constructivism (Learning); *Educational Assessment; Educational Research; Educational Technology; Elementary Secondary Education; *Evaluation Methods; *Junior High School Students; *Learning; Measurement Techniques; Models; Technological Advancement; Test Reliability; Test Validity
IDENTIFIERS *Virtual Reality

ABSTRACT

Preliminary research on virtual reality (VR) suggests that this technology could be a powerful tool for education based on its immersive and dynamic attributes. The Virtual Reality Roving Vehicles (VRRV) Project at the University of Washington is exploring these possibilities by taking VR equipment into elementary and secondary schools for students and teachers to experience and use in building virtual worlds. The question of how to assess VR is particularly significant because of the theoretical conflict between traditional and constructivist learning approaches. This report presents an example of how the VRRV Project is using VR in schools and identifies significant factors for assessment. Junior high school students (n=120) integrated the building of virtual worlds into a curriculum on wetland ecology. With regard to this project and other VR uses, the issue of test reliability versus validity is addressed in terms of general education and using VR. The underlying psychological theories of information processing and constructivism are discussed in terms of developing a comprehensive paradigm to guide the application and research of VR. This discussion is followed by an overview of specific approaches for measuring learning in VR, along with hints and cautions about conducting educational assessment. (Contains 2 figures and 70 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 392 826

Assessing Learning in VR: Towards Developing a Paradigm

Virtual Reality Roving Vehicles (VRRV) Project*

Howard Rose

hrose@hitl.washington.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

HOWARD ROSE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

February 16, 1995

Human Interface Technology Laboratory

University of Washington

all rights reserved

* Prepared while working at the Human Interface Technology Laboratory with generous support from the U S WEST Foundation. The contents of this report are solely the responsibility of the author.

TM024492

BEST COPY AVAILABLE

CONTENTS

I. Introduction: Bringing VR into Schools 1

II. The VRRV approach to assessing learning 3

1. Instructional factors
2. Virtual environment experience factors
3. External factors

III. The value of authentic assessment: Validity Vs Reliability 8

IV. Developing a theoretical paradigm for VR 11

V. Conducting authentic assessment of VR 19

Problem solving
Concept mapping
Metacognitive strategies
Interview techniques
Data gathering
Reciprocal teaching
Computer-based assessment
The effect of VR on other behavior

VI. Analyzing performance 29

VII. Threats to validity and reliability 30

VIII. Conclusion 34

Abstract

Preliminary research of virtual reality suggests that this technology could be a powerful tool for education based on its immersive and dynamic attributes. The Virtual Reality Roving Vehicles (VRRV) Project at the University of Washington is exploring these possibilities by taking virtual reality equipment into schools for students and teachers to experience, and build worlds. Determining the educational efficacy of VR requires developing appropriate and meaningful forms of assessing this new mode of learning. The question of how to assess VR is particularly significant because it exemplifies the broader, theoretical conflict between traditional and constructivist learning approaches.

This report presents an example of how the VRRV Project is using VR in schools, and identifies significant factors for assessment. The issue of test reliability versus validity is addressed both in terms of general education, and specifically in using VR. The underlying psychological theories of information processing and constructivism are discussed in terms of developing a comprehensive paradigm to guide the application and research of VR. This discussion is followed by an overview of specific approaches for measuring learning in VR, along with hints and cautions about conducting educational assessment.

1. INTRODUCTION: Bringing VR into Schools

The Virtual Reality Roving Vehicles (VRRV) Project takes VR technology into public elementary, junior high and high schools and puts it in the hands of students and teachers. Our goal is to evaluate VR as a tool for students to develop broad-based abilities including, but not limited to: problem solving, building mental models, developing effective meta-cognitive strategies and visualization. The VRRV is applying a 'constructivist' approach to instruction which puts each student in charge of their own process of learning. In the constructivist model, the teacher's role is to "support the constructive activities of the learning so that [students'] efforts at constructing understanding--using our cognitive tools--become transparent or ready-at-hand." (Winograd and Flores 1986). Our research mission is to test VR as a medium for making the teaching process "transparent", so students can focus on content rather than falter with the mechanics of instruction.

It is important to ground the discussion of assessment to the VRRV's process of introducing this technology into schools. Before moving ahead, let us look at a sample scenario of how VR is being implemented for this research. In November 1994, the VRRV undertook a month-long world building project with 120 junior high school students at Kellogg Middle School in Shoreline, Washington. The Kellogg Project integrated the building of virtual worlds into a specially designed curriculum about wetlands ecology. Four classes of thirty students participated; each one was randomly assigned to focus on one of the wetlands life cycles: water, carbon, energy and nitrogen. Students learned the fundamentals of their respective cycle according to a constructivist curriculum designed by Kellogg teachers. Each class was then divided into three working groups who each planned and designed a virtual world to express their understanding of the wetlands cycle they studied.

The contributions of the three working groups in each class were brought together and a single virtual world was constructed for each of the four life cycles. The virtual wetlands worlds were populated with plants, animals, objects and landscapes which students created on desktop computers using 3D modeling software. As the final step of the learning process, students put on a VR head mounted display and experienced two of the wetlands worlds, their own plus one other.

The Nitrogen-Cycle World was the most complicated of the four. In this world, students physically manipulated objects in the virtual world and acted out the cycle of nitrification and denitrification as it occurs in a wetlands. Students took free nitrogen, represented by a yellow ball, and placed it in a lightening cloud to demonstrate one way nitrogen is fixed in the atmosphere. The nitrogen then transformed into a fixed nitrogen molecule, represented in the virtual world as a yellow ball orbited by four smaller balls.

Students then flew down to the surface of the wetlands and crossed free-nitrogen with a nitrifying bacteria to fix nitrogen into the soil. The fixed nitrogen emerged within a patch of duckweed to signify the next step in the cycle. The student then picked up a nearby duck and touched it to ("fed it") the duckweed. Immediately, duck droppings and a dead duck appeared on the wetlands shore to indicate the next step along the path for the nitrogen. A denitrifying bacteria (blue ball) also appeared for the student to contact with the decaying matter and release free-nitrogen back into the system to start the process all over again.

As this scenario describes, the process of incorporating VR into the school environment is highly complex and involves human, instructional, and environmental factors . Unraveling these interwoven factors poses a challenge for conducting assessment. A cohesive paradigm to guide

assessment does not exist at this time: One must be created from existing theories of educational assessment, human-computer interaction, and psychology. Considering the substantial financial and human resource investment which may be required to implement VR in schools, comprehensive and accurate assessment of its virtues and weaknesses is crucial in defining the proper role for this technology. This report endeavors to define some parameters and methods for assessing learning with VR, towards the goal of creating a solid theoretical foundation to guide future research and implementation.

2. THE VRRV APPROACH TO ASSESSING LEARNING

The question of how to assess learning using VR is significant because it establishes a scale of relative efficacy for the technology, and also sets the role VR will play in the overall context of education. Preliminary research at the Human Interface Technology Laboratory at the University of Washington (Bricken and Byrne, 1993) and elsewhere (Loftin, Engelberg, & Benedetti, 1993; Regian & Shebilske, 1992; Moshell & Hughes, 1994) gives us an intuitive sense that VR could be highly useful to promote skills and knowledge which students can apply across many domains. The interactive and immersive qualities of VR suggest the potential for an entirely new form of experiential learning.

The instructional model which designates students as passive recipients of declarative knowledge presented in tidy packets has been widely criticized for yielding fragmented and unintegrated learning. Instruction or assessment which is too narrowly focused cannot see the forest for the trees. Glaser (1990) expresses how such fragmentation is especially pronounced in higher cognitive areas such as problem solving.

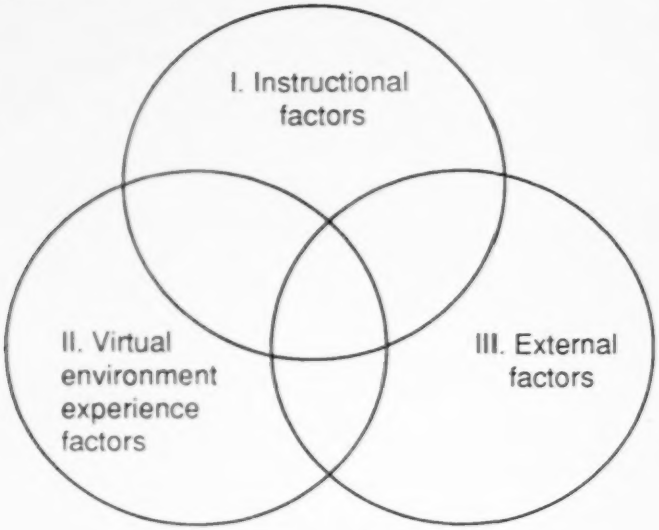
> The danger of fragmentation is that an isolated focus on certain aspects of performance may underlie the frequent findings that students can solve problems but have little ability to explain the underlying principles and that those who can recite or even explain the principles are sometimes unable to recognize the conditions of applicability or to manage the requisite procedures efficiently. A major instructional research task is to design programs that test approaches to the integration of competent performance, and perhaps the most successful approach will be able to test a mix of instructional principles....Attempts at integration promise to provide new grounds for the development of a more encompassing theory of learning.
>
> (Glaser, 1990, p.37)

VR may perhaps give us the opportunity for robust integration, but we must first address the difficult tasks of defining the range of competent performance, and developing assessment methods to adequately measure that performance.

The newness and breadth of the topic of VR can present an obstacle to discussion. Ackerman (1994) describes five leverage points as a basis for discussion and research of VR in education. Her five points include: **transformation** as the world reacts to actions by the user, the qualities of **immersion and point of view**, issues of realism or **verisimilitude**, the sensual engagement of **perceptual and symbolic modalities**, and the factor of **locus of control**. While these points are all important, Ackerman's distinctions still mix together factors of instruction with factors of learning which is inconvenient for discussion of assessment.

For the purpose of the VRRV Project, we have broken our analysis into three categories for assessment: (I) instructional factors, (II) virtual environment experience factors, and (III) external factors. Certain aspects of each of these categories are certain to affect each other (*figure 1*): This interplay must be addressed in order to assess efficacy under real world conditions.

Figure 1 : Assessment Factors



I. Instructional factors

A major research objective is to determine how instruction leading up to and accompanying the students' VR experience influences learning outcomes. Assessment of instructional factors looks at how all aspects of the learning environment *outside* of the head mounted display affect the learning process. Instruction during the world building process, which takes place almost entirely outside of the virtual environment, is one major focus of assessment.

The process of building virtual worlds exemplifies the constructivist paradigm of knowledge being formed within the individual through interaction with the world. Rather than passively receiving information, students can use VR to construct their understanding of the knowledge domain. When children build virtual worlds they are simultaneously structuring their own mental models. Therefore the objects and interactions contained within the world are a direct reflection of the learners' mental models and symbolic representations. Assessment of the world-building

process should take account of how students develop their understanding of the content, how understanding is manifest in the world, and also the quality of the final product.

In the above example of the Nitrogen World, instructional variables include the approach to teaching the background knowledge on wetlands cycles which prepared students to build their worlds, the teaching during the world building process, and the level of guidance which students received as they acted out the nitrogen cycle.

II. Virtual environment experience factors

This category includes the students' experiences and activities while immersed in a virtual world. VRRV assessment focuses on the quality of human-computer interaction, the educational efficacy of various hardware and software interfaces, comparison of world designs, and the physical sensation of presence. In the case of the Kellogg project, the worlds could have been created using different objects, types of interaction, or forms of instruction built into the world. How will such changes to the interface and experience of VR affect learning outcomes?

"The experience in which an idea is embedded is critical to the individual's understanding of and ability to use that idea." (Duffy & Jonassen, 1992, p. 4) In other words, experience is a vehicle for knowledge creation and also recall. Students can experience VR to build their understanding from the ground up. Winn (1993) suggests that VR can give students a physical and intuitive understanding of abstract concepts prior to tackling symbolic representations of the domain. The key to developing intuitive understanding lies in the interactive nature of VR, but care must be taken to avoid misconceptions based on incorrect intuition.

Our research targets a number of important questions regarding how different forms of interaction impact the quality of learning in VR. How do a broad age range of children respond to virtual interfaces? How much learner control of the virtual environment is optimal? If guidance is to be given to the student, should it take place in the virtual environment using an avatar or animated guide, for example? Taking the example of the Nitrogen-Cycle World, was it the physical act of placing nitrogen in a cloud which helped students understand and remember the concept, or would a passive experience of the interaction be equally as effective?

Another assessment area examines the effect of various forms of feedback to support and guide the user. How should a virtual world react to student interactions? Winn (1987, 1992, 1993) and Winn and Bricken (1992) suggest the importance of dynamic feedback in virtual worlds to support learning. Winn (1992) suggests that virtual worlds can be imbued with the ability to support students construction of meaning. Thus it is important to study the relative effectiveness of various modes of feedback. In addition, the level at which students rely on feedback can also be an assessment measure of performance. In other words, the more competence a student develops as she moves from novice to expert within a content domain, the less the student will rely on feedback for guidance.

Winn (1993) suggests that the greatest educational benefit of VR is its spatial qualities of being immersed in another reality. This feeling has come to be referred to as presence by VR researchers, even though a clear method for establishing levels of presence is yet to be established (Hoffman, Hullfish, & Houston, in press). Held and Durlach (1992) propose that synthetic, computer generated environments might enhance the performance of humans operating remote robots. Sheridan (1992) speculates that presence

may improve sensori-motor or cognitive performance. While little is currently known about the phenomenon of presence, VRRV research is delving deeper into the potential benefits of immersion.

III. External factors

Numerous factors unrelated to the VR technology itself will undoubtedly have a crucial impact on students' learning achievement. These factors include differences in individual classroom environments, student characteristics such as personal history or attitudes towards computers, teachers' attitudes and background in technology, and an assortment of social, economic and political variables related to schools, education and technology. A comprehensive assessment of VR technology must take account of how these external factors contribute to the overall context in which VR is applied.

3. THE VALUE OF AUTHENTIC ASSESSMENT: VALIDITY vs. RELIABILITY

The challenge of assessing learning goes beyond determining the efficacy of a single technology: Assessment is inseparable from the broad goals of education. Scholastic measures which do not match classroom teaching lock students in a no-win situation. Measures must be valid and meaningful reflections of skills and knowledge that students can transfer from classroom to the world outside school. Meaningful assessment reflects meaningful instruction.

A major rethinking of educational assessment has begun across the United States. Forty states are in the process of enacting legislation or developing new assessment standards (Pipho, 1992, cited in Taylor, 1994, p.234) . We must consider the evaluation of VR in the broad context of this educational reform. The new wave of standards includes performance

measures such as short-answer questions and student portfolios (Taylor, 1994). Thus we also must develop new rubrics of educational efficacy which illuminate how VR can best fit into the new educational landscape.

Traditional assessment has overemphasized test reliability at the expense of validity (Taylor, 1994; Moss, 1992; Linn, Baker & Dunbar, 1991; Wiggins, 1989). Measures of learning, particularly achievement tests, have almost exclusively been multiple-choice tests of declarative knowledge. Priority in testing has been given to test administration and reliability for reasons of convenience to the testers, but at the cost of students (Taylor, 1994; Sternberg, in press). The result is that current testing procedures give us little meaningful information about what children are learning and are capable of doing (Linn, Baker & Dunbar, 1991). This testing paradox is evident at every level of compulsory education, expressed in textbooks, curriculum and tests.

Breaking free from this paradox will require changing both assessment practices and the content of curriculum. School experiences often fail to match the expectations of the real-world (Duffy & Jonassen, 1992). Numerous researchers (Resnick, 1987; Brown, Collins, and Duguid, 1989; Sherwood, Kinzer, Hasselbring, and Bransford, 1987) have pointed to these disparities as a major underlying cause of failure to transfer school-based learning.

Traditional testing requires numerous inauthentic constraints as indirect proxies for performance to preserve validity (Wiggins, 1992). Typical artificial constraints include: access to reference materials, time restrictions, or limits to the prior knowledge of tasks and how they will be assessed. Constructivists, such as Jonassen and Duffy, attack such artificial testing constraints as ineffective techniques for measuring what is significant about student abilities. They believe "the critical aspect of performance is the ability to respond to the situation constraints - to be able to construct new plans

based on the changing demands and constraints of the situation." (Duffy & Jonassen, 1992, p. 4) Thus testing in the constructivist paradigm is carried out in the closest approximation of the real-world performance environment as possible. Wiggins offers an interesting example of a more appropriate testing constraint: A physics teacher allows students to bring an index card to the exam with whatever notes they choose. The teacher collects the cards after the test, and notes that the content of the cards often reveals more about the students' knowledge than the exam answers (Wiggins, 1992, p. 31).

The growing popularity of authentic assessment is pushing the development of measures which are valid reflections of students' ability and knowledge. However, authentic assessment does not merely mean using new methods to measure the same old learning. In his critique of science assessment, Shavelson, Baxter, & Pine (1991, p. 355) notes how performance-assessment approaches measure something significantly different about the scientific process than do traditional multiple choice tests. Instead of testing retention of verbal information, constructivist assessment tests the presence of more general indicators of learning such as mental models or the ability to construct plausible solutions to previously unencountered tasks. Cunningham (1992, p. 42) explains: "We check to see if the student is developing self-awareness of the constructive process: the context-specific nature of interpretations, the value of multiple perspectives, the relativity of positions, etc." Constructivist assessment is often embedded in the learning process.

Authentic assessment approaches have been criticized on the grounds that they are not reliable and are difficult to generalize across student populations. Some of these criticisms and possible solutions appear below.

This discussion of general trends in educational assessment is significant because it suggests a growing need to widely adopt performance assessment. Thus the assessment standards and methods chosen for VR must match with the broadly accepted practice in schools. Conversely, VR may offer a highly controllable testbed to enhance the quality and reliability of performance assessment. The power of VR as a tool for both experiencing prebuilt worlds and, more importantly, world building by students, suggests the technology will be widely applicable for education. It is crucial to consider VR performance assessment within the general context of authentic assessment because VR developers need to anticipate the overall educational environment in which the technology is to play a role.

4. DEVELOPING A THEORETICAL PARADIGM FOR VR IN EDUCATION

Because the theory underlying the design of assessment tasks inevitably shapes the final form of assessment, it is essential to clarify the theoretical basis for assessment from the outset. Further research and application of VR will benefit from a well developed and appropriate working paradigm for applying the technology in education.

The information processing model of human cognition has long been the predominant paradigm in psychology, human-computer research, educational research and the field of assessment. Information processing has been heavily influenced by the computational model of cognition (Newell & Simon, 1972), especially in the study of human-computer interaction. According to information processing paradigm as stated by Lachman, Lachman and Butterfield (1979, p. 99), cognitive psychology and computers share a lot in common. "It [cognitive psychology] is about how people take in information, how they recode and remember it, how they make decisions,

how they transform their internal knowledge states, and how they translate these states into behavioral outputs." This paradigm stands firmly rooted in the objectivist tradition.

Other information processing researchers such as Anderson (1983, 1990) have enhanced the computational model to make it more relevant to education. Anderson's theory of Adaptive Control of Thought (ACT*) moderates the information processing model to make it more applicable to describe learning. ACT* has enjoyed rather wide acceptance, yet ACT* does not address some of the key elements of learning deemed important in the constructivist paradigm such as student motivation and attitude. Nor is current information processing theory robust enough to describe highly complex, integrated learning as it often happens in the real world.

Jonassen (1992, p.138) charts the theoretical ideals of objectivism and constructivism as polar opposites. He notes, however, that in reality instructional designers tend to fall somewhere in the middle of this continuum.

objectivism <-------PI---------ID----------ITS--------Piagetian------>constructivism
externally mediated reality internally mediated reality

(PI: programmed instruction; ID: instructional design; ITS: intelligent tutoring systems)

The conflict over the validity of the objectivist approach to instruction and learning assessment is at the crux of what sets these two approaches apart. Is the act of learning merely the completion of a set of processes, as information processing suggests? Or is learning the act of constructing parts into a greater, more meaningful whole? A complete assessment of the educational efficacy of VR requires supplementing the useful aspects of both the information processing and constructivist approaches. Following are

brief descriptions of the two paradigms. The purpose is to suggest what aspects of information processing may be appropriate to our assessment, and to clarify the unique aspects of constructivist assessment.

4.1 *Information Processing*

A main feature of the information processing approach is the emphasis on a well defined understanding of expert behavior. The target knowledge domain is established from the outset and assessment is based on how closely a novice student is able to approximate the competence of an expert. Competence as described by Glaser (1990, p. 30) has three major aspects: "(a) the compiled, automated, functional and proceduralized knowledge characteristic of a well-developed cognitive skill; (b) the effective use of internalized self-regulation control strategies for fostering comprehension; and (c) the structuring of knowledge for explanation and problem solving."

Anderson's (1983) ACT* model has been widely applied to computer-based training. The ACT* model is particularly relevant to learning assessment in VR because of its focus on higher cognitive skills. Anderson (1983) names three stages to describe the transition from novice to expert.

> Declarative Stage: knowledge is stored as bits of declarative information
>
> Knowledge Compilation Stage: Transition of verbal information to more complete mastery, or skill level. This stage features
>
> > *Composition*: Combining sets of steps into single steps which can be executed easily;
> >
> > *Proceduralization* developing condition/action responses to stimulus or situations.

Procedural Stage: Streamlining the set of procedures and strengthening the processes.

The ACT* paradigm calls for a cognitive task analysis for each task before training and testing the skill.

Royer, Cisero, & Carlo (1993) published a survey of techniques for assessing higher cognitive skills based on the paradigm of Anderson's ACT* model. Their approach breaks information processing into three distinct layers: 1) basic capacities; 2) cognitive skills capable of being transformed from controlled to automatic/encapsulated processes; and 3) higher cognitive skills for goal setting and planning cognitive activity. Assessment at any of these layers requires determining the current stage of skill development, not simply if a certain skill has or has not been acquired. Royer, Cisero, & Carlo (1993, p. 207) also suggest a helpful framework for categorizing cognitive skill assessment techniques:

Knowledge organization and structure: Storage as loosely related facts. Measure of knowledge organization and structure development is an indicator of higher cognitive skill.

Depth of problem representation: Perception of the problem as abstract principles. The novice perceives problems in terms of particular elements, not as a generalized set. The ability to perceive the principles underlying a problem is an index of skill development.

Quality of mental models: The ability to imagine a system in operation. The model guides performance working within the domain. The presence and sophistication of mental models is a measure of skill development.

Efficiency of procedures: Eliminating unnecessary steps in solving a problem. The ability to efficiently use acquired skills is another index of growing skill development.

Automaticity of performance: Efficient handling of cognitive load leaves room for extra processing of integrating information. Assessment tasks should systematically represent the critical performing a completely unrelated task. Automatic and capacity-free performance is a measure of skill development.

Metacognitive skills: Ability to reflect on and control performance efficiently. The ability to plan activity, monitor outcomes and alter behavior accordingly demonstrates skill development.

Figure 2 (Royer, Cisero, & Carlo, p. 1993, p. 209-10) is helpful for matching specific task types to target cognitive dimensions See Royer, Cisero and Carlo's text for a detailed explanation of each task.

While the information processing paradigm offers a strong basis to analyze human-computer interaction, it is important to acknowledge that there are other paradigms through which to make assessment. In light of the weakness of current information processing theory to guide research in the creation of complex, integrated learning environments and to take factors such as attitude and motivation into account, assessment of educational VR would seemingly benefit from a broader and more robust paradigm of learning.

figure 2: from Royer, Cisero, & Carlo, p. 1993, pp. 209-10.

Cognitive Skill Assessment Techniques
Cognitive Dimension Assessed

Author	Type of Task	Development Level of Cognitive Skill
Knowledge Acquisition		
	Traditional assessment	
Ronan et al, 1976	Fireman tab test	Declarative
Lesgold & Lajoie, 1991	Recall of electronic components	Declarative
Knowledge Structure and Organization		
Shepard, 1962	Multidimensional scaling	All levels
Geeslin & Shavelson, 1975	Associative recall of concepts	All levels
Chi et al, 1982	Conceptual recall of physics concepts	All levels
Konold & Bates, 1982	Concept ratings	All levels
Konold & Bates, 1982	Concept categorization	All levels
Reitman & Rueter, 1980	Concept free recall	All levels
Adelson, 1981	Free recall of computer programs	All levels
Gutherie, 1988	Document search	All levels
Card et al, 1980	Text editing	All levels
Royer, 1990	SVT assessment	All levels
Carlo et al, 1992	Inferencing assessment	All levels
Depth of Problem Representation		
Chase & Simon, 1973	Chess perceptual reproduction	All levels
Chase & Simon, 1973	Chess memory reproduction	All levels
Egan & Schwartz, 1979	Reproduction of electronic circuits	All levels
Barfield, 1986	Program recall	All levels
Chi et al, 1981	Physics problem sorting	All levels
Schoenfeld & Hermann, 1982	Math problem judgments	All levels
Carlo et al, 1992	Classification of scientific principles	All levels
Adelson, 1984	Flowchart comprehension	All levels
Adelson, 1984	Insert missing line of program code	All levels
Goulet et al, 1989	Identification of tennis serves	All levels
Allard et al, 1980	Recall of basketball positions	All levels
Purkitt & Dyson, 1988	Information usage in political decision making	All levels

figure 2 continued:

Author	Type of Task	Development Level of Cognitive Skill
Mental Models		
McClosky et al, 1980	Prediction of flight path	Declarative/Compilation
Gentner & Gentner, 1983	Identifying underlying metaphors	Declarative/Compilation
Lopes, 1976	Poker mental models	All levels
J.R. Anderson, 1990	Correct and buggy productions	All levels
Johnson, 1988	Malfunctioning generator models	All levels
Lesgold et al, 1988	X-ray drawing	All levels
Metacognitive Skills		
Baker, 1989	Text faulting	All levels
Rosenbaum, 1986	Visit planning	All levels
Gerace & Mestre, 1990	Planning in physics problem solving	All levels
Lesgold et al, 1990	Problem space planning	All levels
Sweller et al, 1983	Changes in problem solving strategy	All levels
Automaticity/Encapsulation of Performance		
Lesgold & Lajoie, 1991	Speed of conceptual processing	All levels
Schneider, 1985	Dual task methodology	All levels
Britton & Tesser, 1982	Dual task methodology	All levels
Efficiency of Procedures		
Glaser et al, 1985	Card sorting of assembly procedures	All levels
Lesgold & Lajoie, 1991	Multimeter judgment	All levels
Lesgold & Lajoie, 1991	Multimeter placement	All levels
Lesgold & Lajoie, 1991	Logic gate efficiency	All levels
Green & Jackson, 1976	Hark-back technique	All levels

4.2 *Constructivism*

At this time, the question of how to assess learning in the constructivist paradigm has gone largely unaddressed. Jonassen is one of the few who has attempted to outline what constructivist assessment might look like.

> As evaluators we need to focus on learning outcomes that will reflect the intellectual processes of knowledge construction. Clearly, knowledge construction entails higher order thinking. So, outcomes of constructivistic environments should assess higher order thinking, such as that at the "find" level of Merrill's (1983) taxonomy, the "cognitive strategy" level of Gagne's (1987), and the "synthesis" level of Bloom's taxonomy.
>
> (Jonassen, 1992, pp. 140-1).

Thus assessment of learning in the constructivist paradigm can perhaps be evaluated with modified versions of existing taxonomies and strategies. Whatever methodology is chosen, it is clear that assessment must address both the process of knowledge acquisition as well as the final product. Toward this end, constructivists propose embedding assessment in the actual learning process. To do so is in sharp contrast to teaching and evaluation approaches which only test cumulative skills and knowledge after the learning process has been theoretically completed.

Based on the constructivist conception that learning is an individualistic endeavor, Jonassen (1992) suggests that each individual learner may be the only one capable of interpreting his or her own progress. Therefore Jonassen believes that the evaluation of learning should be goal free relative to external criteria of success. But he also recognizes that constructivism needs to develop valid methodologies for assessment in order to gain wider acceptance. Jonassen cites Scriven (1973) for proposing needs-based assessment methods as the most objective standards by which to

evaluate outcomes of any process. "Criterion-referenced instruction--where the goals of learning drive the instruction--and evaluation are prototypic objectivistic constructs and therefore not appropriate evaluation methodologies for constructivistic environments." (Jonassen, 1992, p.140)

Authentic tasks must be relevant to the real world relevance and utility of learning and should integrate knowledge across subject areas. "Simplified, decontextualized problems are inappropriate outcomes for constructivistic environments. So are they for evaluation, as well." (Jonassen, 1992, p. 141). Jonassen offers some specific suggestions to describe--even if in only very sketchy, embryonic terms--characteristics of desirable assessment.

- "Rather than learning being referenced by a single behavior or set of behaviors, it should be referenced by a domain of possible outcomes, each of which would provide acceptable evidence of learning."
- Should have a panel of reviewers, each with a meaningful perspective and reasonable credentials.
- A novice might provide a better evaluation than an expert, who frequently focuses on inappropriate criteria of learning.
- Evaluation of multiple products or outcomes is preferable to assessing only a single one.
- "Evaluation from a constructivistic perspective should be less of a reinforcement and/or behavior control tool and more of a self-analysis and metacognitive tool."

(excerpted from Jonassen, 1992, pp. 143-5)

General agreement is yet to reached on what types of knowledge domains are appropriate for constructivist teaching. Jonassen (1992) suggests that constructivistic learning environments are most appropriate for advanced knowledge acquisition, while it is likely that introductory knowledge acquisition is better supported by more objectivistic approaches. Fosnot (1992, p.172) is critical of Jonassen's position. "In my mind, he

[Jonassen] has missed the main point of constructivism. Learners are always making meaning, no matter what level of understanding they are on. Constructivism is not a theory to explain only complex, ill-structured domains; it is a theory of how learners make meaning, period!...To assume the learner is a blank slate until presented with information, and to characterize experiences or tasks separate from the learner's meaning of them, is objectivistic--a perspective which in the first chapter Jonassen (& Duffy) so radically opposed!" Winn (1992, p. 179) expresses "I am not yet convinced that all knowledge can be constructed by students. The student has to have some knowledge from which to start construction. And that knowledge needs to be explicitly taught. Constructivists may well disagree with this."

In summary, the constructivist paradigm differs from information processing in a number of fundamental ways. Unlike information processing, constructivism considers factors of motivation and interest to be crucial to the learning process. Constructivism stresses integration of diverse knowledge, rather than reducing the complex "behaviors" of experts into subroutines. In terms of tasks for assessment, while information processing tasks are very often performance based, the tasks are defined for the student in very specific ways. Constructivist tasks are student centered -- often student generated -- and can result in a wide assortment of possible responses.

VR may prove to be an optimal media for conducting constructivist assessment as well as instruction. The dynamic nature of the computer system allows recording of student interactions and data gathering in the background as the student moves through the virtual world. Once recorded, the record can be reviewed by the student to reconstruct and evaluate the learning process. Thus the application of VR as an assessment tool, in and of itself, is another promising area for research.

5. CONDUCTING AUTHENTIC ASSESSMENT OF VR

It is a common practice of authentic assessment to embed the test instrument into the learning process (Wiggins, 1989, 1992; Linn, Baker & Dunbar, 1991). Wiggins (1992) states that good assessment *is* good instruction. This point is crucial because it implies that the factors which contribute to good instruction are *themselves* the measurement tool for assessment. One example is the earlier mention of offering constructive feedback to the learner. The quality of feedback will influence learning. At the same time, student reliance on feedback can be interpreted as an indication of competence. This inter-relationship cannot be ignored when establishing assessment criteria and measures.

When writing test questions, the questions themselves can serve as exemplars of good teaching practices that are not likely to distort the teaching and learning process. Linn, Baker & Dunbar (1991, p. 16) suggest that questions should not be directly teachable; however, teaching for them will result in good instruction. Understanding the basis on which performance will be judged also promotes improved performance.

Below is a list which includes a range of authentic assessment methods and approaches. Since it is beyond the scope of this paper to give in-depth discussions of the merits and virtues of each, references have been included for each category to direct the reader to relevant sources.

5.1 *Problem solving*

Problem solving involves complex interactions between a multitude of cognitive, metacognitive and knowledge-based processes. Szetela and Nicol

(1992, pp. 43-4) break the problem solving process down into three stages: a) understanding the problem; b) solving the problem
c) answering the question, and score performance on each one separately. This presents a more detailed picture of students' abilities than a simplistic approach such as measuring only correct and incorrect outcomes. Szetela and Nicol also identify the following typical sequence of actions for successful problem solving:

1. Obtain appropriate representation of the problem situation
2. Consider potentially appropriate strategies
3. Select and implement a promising solution strategy.
4. Monitor the implementation with respect to problem conditions and goals.
5. Obtain and communicate the desired goals.
6. Evaluate the adequacy and reasonableness of the solution.
7. If the solution is judged faulty or inadequate, refine the problem representation and proceed with a new strategy or search for procedural or conceptual errors.

When we consider these steps in terms of the characteristics of VR, a clear picture begins to emerge of how VR could aid student problem solving. Let us look at how VR matches with each of the above steps. 1) VR may prove to be a powerful visualization tool for representing abstract problem situations. 2) Virtual worlds allow for a high degree of trial and error, which may encourage students to explore a greater range of possible solutions. 3) The student is free to interact directly with virtual objects which allows for firsthand hypothesis testing. 4) The virtual world can be programmed to offer feedback which focus the student's attention on specific mistakes, thereby enhancing students' ability to monitor their own progress. 5) The VR system can collect and display complex data in real time, which may help students obtain their desired goals. 6) The immersive nature of VR might enhance

students' capability to retain and recall information, which could facilitate the evaluation of solutions. 7) The virtual world is a fluid environment well suited for the iterative process of refinement.

But the question remains as to how to evaluate students' progress along the steps presented above. Szetela and Nicol suggest six approaches for generating questions to stimulate and assess problem solving which are highly applicable to VR: (a) present a problem with all the facts and conditions, but have the students write an appropriate question, solve the completed problem and write their perceptions about the adequacy of the solution; (b) present a problem with a partial solution; (c) present a problem with unrelated facts, have students revise problem; (d) have students explain how they would solve a problem using only words, then do it; (e) after students solve a problem have them write a new one with different context but preserving the original structure; and (f) present a problem without numerals. Students supply numbers, estimate answers and solve the problem themselves.

Another assessment approach might be to have the students create their own evaluation method for worlds they have built. In other words, have students define the learning task and the criteria they would use to evaluate an individual's performance in their world. This process would require students to analyze what information is crucial in their worlds, and to generate their own problems which users would have to solve.

5.2 *Concept mapping*

Concept mapping is a process where students organize a domain of knowledge for themselves and express their understanding of the various inter-relationships in the form of a diagram (Novak & Gowin, 1984). Because

there are numerous ways to diagram any complex set of relationships there is no single "right" answer, making concept mapping an ideal instrument for authentic assessment. The change seen in students' maps from pre-treatment to post-treatment measures their learning and the sophistication of mental structures.

Some educators view story maps as props which should be withdrawn as soon as possible; others see them as useful planning tools in preparation for synthesis activities (Quellmalz, 1991, p. 324). Typical criteria to assess the relative quality of concept maps include the appropriateness of the map to the content, content categories included in the map, the amount and quality of information portrayed, and the level of knowledge organization demonstrated.

The example of the Nitrogen-Cycle World could be judged as a concept map, portraying the student's perception of relationships and processes in the cycle. Students develop an internal concept map during the world building process. Then they must figure out how to express their knowledge to others through the medium of the virtual world. While the technological complexity of VR may hamper students' ability with the medium, there is also a strong possibility for VR to open up a new avenue of innovation and expression.

5.3 *Metacognitive strategies*

There is substantial evidence which links the quality of metacognitive processing with development of knowledge structures (Butterfield, Albertson, & Johnston, 1993). Metacognitive components such as planning, self-monitoring, evaluation and reflection are assumed to be indicators of how closely students approximate the behavior or experts. Quellmalz (1991, p. 322)

uses a technique of having students give reflective accounts to explain what they have learned. The sophistication of the explanation indicates the development of knowledge formation. Another externally visible indicator of metacognition is the students' reliance on feedback and support while using an instructional program, i.e. in the virtual world. The term 'scaffolding' refers to the forms of assistance students require as they progress through the learning process. Scoring rubrics focus on the amount and nature of assistance required (Quellmalz, 1991, p. 324).

5.4 *Cooperative learning*

There is general consensus that students working in small groups produce higher achievement that students working alone, especially in a cooperative setting (Johnson , Johnson, & Stanne, 1985; Yager, Johnson, & Johnson, 1985). The optimum size seems to be either two or three (Cox & Berger, 1985; Webb, Ender, & Lewis, 1986). There is also general consensus that paired students should be like-gendered and have similar abilities (Dalton, 1990; Dalton, Hannafin, & Hooper 1989; Johnson , Johnson & Stanne, 1985 Johnson , Johnson & Stanne, 1986).

A common conception of VR, and computer technology in general, is that it isolates the user and reduces human interaction. One of the stated missions of the VRRV project is to explore how VR can be used to enhance human interactions in a number of contexts. First, there are many opportunities to encourage group collaboration within the design phase of world-building. Second, the experience of a single student in VR does not have to be conducted in isolation. Possibilities include interactions between a student immersed in a virtual world and those outside, or the interaction between students watching another using VR. Finally, the VRRV Project has

the technological capability for two students to share the same virtual space and collaborate on a single task. While the a review of the literature on collaborative learning effects is beyond the scope of this report, I would like to mention two relevant studies of the educational effects of collaboration in computer-based training.

Stephenson's (1991) study of computer-based training found that students benefited from teacher-student interaction of a social nature, and also through paired-learning arrangements. He also concluded that the relationship between students took the place of teacher-student interaction, since the most successful students were those who were in paired groups, followed by individuals who had high teacher-student interaction. Stephenson also found that weak students are more impacted by lack of social interaction than are strong students. These findings indicate that the one-student: one-computer model of computer-based training may be essentially flawed because it negates the social aspects of learning.

Dalton (1990) found that it is not merely the presence of collaboration which contributes to learning, but the quality of the interactions which is the determining factor. He found that structured learner interactions aid encoding and cognitive process, and high-level elaboration (where students explain the content out loud) is the critical, beneficial factor of collaboration. Thus assessment of VR must measure more than the frequency of interaction; it must measure the propensity of VR to stimulate meaningful and productive collaboration.

These studies suggest that the VR technology which fosters collaboration will yield even greater educational benefits. The question for research then becomes how to encourage meaningful collaboration both inside and outside virtual space? Attention must also be given to how to

train instructors to promote desirable interactions when using VR. Interestingly, if one establishes that the quality of student interactions is correlated with learning and performance achievement, then a measure of that quality becomes an indirect method of assessment.

5.5 *Interview techniques*

Interviewing is a central technique for authentic assessment because of the value and emphasis placed on the experience of individual learners. Interviews may be open ended or highly structured depending on the type of assessment and the age of the subjects. In the process of explaining their thinking or learning process, students reveal more than if they can correctly answer test questions. The language and manner in which the student explains herself gives insight into how developed their cognitive models of the domain are. Specific interviewing techniques include using probing questions, having the subject do free association, and video taping student performance then replaying the video while the subject recounts the experience (Suchman & Trigg, 1991).

Role playing exercises can be a revealing element of interview or debriefing sessions. Kourisky (1983) reports facilitating instructor-led, inquiry-oriented discussion and role playing sessions as a means to focus students' attention.

It is important to keep in mind that students may not be able to express their own ability and knowledge accurately to the interviewer. Some students may be better at performing an investigation to solve a problem than they are at verbally explaining the operations involved in an investigation.

5.6 *Gathering data from performance tests in VR*

Some possible data gathering techniques to assess performance in a virtual environment include: video tape and analyze the subject's body movements in VR, observe quality and level of student interaction with the world, monitor the interaction between students watching someone experience VR, and monitor the amount and types of assistance the student requires to perform tasks.

5.7 Reciprocal *teaching*

Brown and Palincsar (1984, 1989; Glaser 1990) describe reciprocal teaching as an instructional procedure where "students take turns in leading the class in the use of strategies for comprehending and remembering text content that the teacher models for the class. Its three major components are (a) instruction and practice with executive strategies--questioning, summarizing, clarifying and predicting in the course of reading text--which enable students to monitor their understanding; (b) provision, initially by a teacher, of an expert model of these metacognitive processes; and (c) a social setting that enables joint negotiation for understanding." In addition to being a successful instructional practice, reciprocal teaching is also an effective device for assessment. As a student organizes and verbalizes her knowledge to teach another, the extent to which their understanding has developed becomes visible. "The Reciprocal Teaching method creates a zone of proximal development where learners perform within their range of competence while being assisted in realizing their potential levels of higher performance (Vygotsky, 1978)." (cited in Glaser, 1990, p.33).

Rosenshine and Meister (1994) have made a comprehensive review of reciprocal teaching research which should prove a useful guide for designing assessment.

5.8 *Conducting computer-based assessment*

In the current context, computer-based assessment refers to conducting assessment using a conventional PC platform to test transference of learning out of the virtual environment. Using flat-screen, computer simulations also offers an alternative computer environment for comparison with VR.

Computer-based assessments have a well established track record and offer some attractive advantages over hands-on or paper-and-pencil testing methods. Automating with computers means assessment is less costly and time consuming to administer compared to hands-on or interview assessments. The computer maintains a full record of performance for easy review of problem solving process. Embedding assessment in a computer program can also offer advantages for the student and boost performance. For example, students can experiment with the technology to discover solutions to problems that are unavailable in other types of assessments.

Nelson et al. (1993) describe methods for using data gathered by the computer as users move through a hypermedia system. Assessment can be based time spent on particular screens, the paths taken as the user moves from node to node within the system, or qualitative evaluation of social interactions matched with the record of human-computer interactions. These techniques apply to assessment of conventional multimedia, and could also be adapted for immersive VR.

A study conducted by Kumar (1994) used a HyperCard stack to assess learning. He found that HyperCard and pen-and-paper assessment methods

influenced the performance of expert and novice students differently in tasks to balance chemical equations. In a test of learning in high school chemistry, Kumar found that students scored significantly higher using a computer than with pen-and-paper. Novices using HyperCard actually did as well as experts with pen-and-paper! Kumar credits the advantage to the computer's ability to remember for the students, which reduces their overall cognitive load. The computer also give immediate feedback which improves motivation and attention to the assessment task. Hypermedia can provide a non-linear environment for problem solving to allow the transfer of knowledge across domains (Kumar 1994, p. 64). Kumar's study is a good illustration of how a test can become a teaching tool.

Some potential dangers in using hypermedia for assessment should be mentioned here. Researchers have found that it can be difficult to keep students on task in large hypermedia systems; students may become disoriented within the program (Kumar 1994); and there may be a gender bias favoring males (Clarke, 1990). For detailed discussion of how and why to use computer based assessment approaches see Shavelson, Baxter, & Pine (1991) and Kumar (1994).

5.9 *The effect of VR on other behavior*

Assessment should not overlook possible residual benefits and changes resulting from the introduction of VR into the classroom. Potential areas for study include: (a) increased use of computers, (b) changes in student self-image and confidence, (c) implications of technology elsewhere in the classroom, and (d) carry over to other areas of student interest.

6. ANALYZING PERFORMANCE

In addition to creating valid tasks, we must also conduct valid analysis of the data. Reeves (1986, 1992) is a sharp critic of the outcome of most experimental and quasi-experimental designs in education. His review of the literature found that few research and evaluation efforts have reported any statistically or educationally significant differences (Reeves, 1986, p. 102). Winn (Winn, 1993) cautions that "...instructional designers are wrong to assume that they can base instructional strategies on the analysis of an objective, standard world... evaluation of learning can only tell us what students appear, or pretend to know, not what they really know." (Winn, 1993).

Reeves (Reeves, 1986, p. 103) suggests the need for a new paradigm of assessment to draw more meaningful conclusions about educational media. His two step approach to monitor the assessment process is as follows:

Step 1: measure differences in:

a) initial characteristics of learners

b) contextual variables

c) dimensions of the instructional treatment

d) criteria or outcomes.

Step 2: Analyze measured differences in terms of:

a) How much variance in outcomes can be uniquely attributed to each of the predictor domains (student initial abilities, context and treatment)

b) How much variance can be attributed to interactions among the predictor domains?

The measurement of cognitive gains via constructing a causal model of critical dimensions of VR which influence learning outcomes is based in the information processing paradigm; the antithesis of constructivism. Reeves suggests basing such a causal analysis on Gagne's (1974) nine events of instruction which is heavily based on the assumptions of the computational model. An attempt to construct such a model may indeed prove helpful in understanding VR, and to ground the study of this new technology in the proven and accepted legacy of the old. It is important to note, however, that such an exercise would mean little when viewed from the constructivist perspective.

7. THREATS TO VALIDITY AND RELIABILITY

Shavelson, Baxter and Pine (1991) examine these criticisms and conclude that authentic assessment approaches can yield reliable results if each hands-on investigation is treated individually, with the obvious disadvantage that such procedures are far more time and labor intensive than traditional paper-and-pencil examinations. Authentic testing methods are also delicate instruments which require fine tuning and great care in administration. Inter-observer consistency is one of the major threats to reliability for many strategies (Kazdin, 1982). Authentic tasks and tests are often extremely heterogeneous: some are more difficult than others and they can vary widely in the specific knowledge-domain which they assess. Test results show that individual student performance can vary dramatically on similar test items and tasks. Many tests may also be biased toward students with previous experience in hands-on learning. Another criticism is that techniques such as self-reporting or interviews rely too heavily on an individual's verbal and communication abilities as an information source.

Perhaps most importantly, Shavelson, Baxter and Pine (1991, p. 32) note that "a substantial number of assessment tasks are needed to generalize, with any degree of confidence, from students observed performances to the science domain of interest."

Educational assessment involves countless factors which could disrupt, alter or invalidate data collection that researchers in the physical sciences never need to address. Some of these problems can be attributed to the nature of working with human subjects, others to the environment of school administration and classrooms. The literature on assessment contains substantial warnings of potential pitfalls which are worthy of noting.

One of the primary concerns in conducting complex assessment is to insure consistency across treatments and the rating of student performance. To guard against inter-observer error, conduct trial assessments using video-examples of sample subject performance to train assessment administrators (Blumberg et al., 1986; Suchman & Trigg, 1991). Administrators should practice with the tape and compare their results until agreement on scoring is reached. Wiggins (1992) suggests developing a detailed protocol of how tasks should be administered to insure that judges will know the proper limits of their interventions to student acts, comments or questions. He notes how easy it is to completely invalidate a study's results with inconsistencies.

If assessment relies on classroom teachers making and recording observations, it is helpful to make tasks maximally self-sustaining and the record-keeping obligation mostly the students'. Systematization and automation of the assessment process will free the teacher to focus on more valuable judgments (Wiggins, 1992).

Ogborn (1994) makes a number of cogent cautions regarding the design and exploration of learning environments. He points out some difficulties in

designing tasks for testing expressive, as opposed to exploratory, use of software. Task goals must be concise and clearly explained to the user. Also, ample time must be allowed so the user progresses beyond mastering the interface to focusing on the content of the task. Ogborn criticizes much research for expecting to achieve learning gains with unrealistically short treatment times. "Most worthwhile learning takes a good long time to achieve, best measured in weeks or months than in days or hours." (Ogborn (1994, p. 35).

Gender bias is one potential confounding factor in educational assessment, particularly in research related to technology. Clarke (1990) advises researchers to take account of external influences which may create gender effects when developing test questions. For example, he found that test questions which involved female-stereotyped activities such as determining the most effective flooring for a kitchen did not engage some boys.

Specific problems may arise in certain domains of knowledge do to students' preconceived notions and attitudes. Clarke (1990) found students' views of what is or is not "science" are shaped by personal experience. Consequently, students may reformulate an assessment task to fit their perception of science and proceed to solve the problem in ways incompatible with those intended.

Researchers must also be cautious of the influence of developmental changes and age specific phenomena on research results. The method in which assessment activities are administrated must be consistent across all age groups to take account of developmental changes in problem-solving. This will also help determine which activities are inappropriate for a given age group.

Another potential source of confounding variables can generally be characterized under the heading of learner types. That is, specific learner characteristics such as prior knowledge, general aptitude, gender, learning style, socio-economic background or previous experience with technology might significantly influence learning with VR for specific students.

While it is beyond the scope of this report to even begin to address the numerous individual differences worthy of study, let us look as the single characteristic of field dependent versus field independent learners as a case in point. A significant number of studies (Frank and Keene, 1993; Davis & Cochran, 1989; Frank, 1983) suggest a significant distinction between field dependent and independent learning styles. The construct of field independence-dependence refers to the stable and pervasive preference of individuals for either analytical or global information processing. Field-independent individuals are strong in perceptual and conceptual tasks, actively segmenting information into relevant parts and analyzing the interrelationships among those parts. Field-dependent individuals process information in a global, holistic, and passive fashion; their processing tends to be dominated by the existing organization of the perceptual and cognitive field (Goodenough, 1976).

Future research in VR might be to examine ways to encourage field-dependent students to use a more active and flexible style of information processing. This training could focus on developing a range of skills including metacognitive awareness, mathemagenic memory strategies (i.e. elaboration, categorization, thematic organization), or incorporate Vygotsky's (1978) concept of the proximal zone of development within cooperative group training activities (Johnson & Johnson, 1987; Slavin, 1986). VR could be a vehicle to encourage active processing strategies for field-independent

students by offering direct, physical interaction and manipulation of abstract content.

8. CONCLUSION

A comprehensive evaluation of the educational efficacy of VR must take account of all three factor areas for assessment: instructional, experiential and external. Meaningful assessment requires robust rubrics and standards in order to illuminate the unique aspects of VR. Student performance with the technology should be observed and rated over an extended period of time and include the learning process, not merely a single test of outcome. Assessment procedures must be relevant to content area. When assessment is embedded in the learning process, it is important to clarify the distinction between individual factors, such as feedback or cooperative learning, which can be both an independent variable of instruction or an assessment measure.

Considering the incomplete nature of the field at this time, the key to conducting meaningful assessment will be to apply multiple measures of learning and performance. Reciprocal teaching and open ended interview techniques will yield the greatest bounty of data, but these methods suffer from being labor intensive and weak at yielding quantifiable comparisons. Perhaps the most promising form of assessment will be to use the computer to capture motions and interactions, which significantly speeds data collection and can also become a basis for students to recount their experiences. A variety of interview techniques such as role playing will enhance the interview process, especially for young children. Well designed instructional software which mimics the virtual world will be good tests of transference, and will also enable automated data collection for assessment.

In the case of assessing the world building process, it may be beneficial for students to formulate their own evaluation methods. The process of stating criteria for successful completion of a worlds, stimulates reasoning and problem solving skills, encourages students to teach and test one another, demonstrates that students grasp fundamental and critical knowledge, and reinforces learning. This practice follows the constructivist paradigm through student centered learning, embedding assessment into the learning process, and allowing for open ended outcomes tailored to individual students.

Tests of complex levels of cognition such as problem solving, building mental models and metacognition will need to be adapted to fit the nature of VR. Tasks must be not only engaging for the students, they must address the unique, immersive nature and interactive aspects of VR so as to distinguish the level of learning directly attributable to the technology. As a general principle, research and development of VR should strive to encourage greater human-human collaboration and interaction, possibly using the level and quality of this interaction as a measure of success.

Research using VR is susceptible to every validity and reliability confound in conventional assessment, plus a whole new set related to the technology. Thoughtful application of theory to practice should reveal the potential.

Bibliography

Ackerman, E. (1994). Direct and mediated experiences: Their role in learning. In R. Lewis and R. Mendelsohn (Eds.), Lessons from learning. Proceedings of the International Federation for Information Processing (IFIP) Working Conference on Lessons From Learning (pp. 13-21).

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. (1990). The adaptive character of thought. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barfield, W. & Weghorst, S. (1993). The sense of presence within virtual environments: A conceptual framework. In G. Salvendy & M.J. Smith (Eds.) Human-Computer interaction: Software and hardware interfaces.

Blumberg, F., Epstein, M., MacDonald, W., & Mullis, I. (1986). A pilot study of higher-order thinking skills assessment techniques in science and mathematics--Part I and Pilot-Tested tasks--Part II. Final Report. Princeton, NJ: National Assessment of Educational Progress.

Bricken, M. & Byrne, C. (1993). Summer students in virtual reality: A pilot study on educational applications in virtual reality technology. In A. Wexelblat (Ed), Virtual reality applications and explorations. Toronto: Academic Press Professional. (pp. 199-217).

Brown, A.S. & Palinscar, A.S. (1985). Reciprocal teaching of comprehension strategies: A natural history of one program for enhancing learning. (Tech. Rep. No. 334). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Brown, A.S. & Palinscar, A.S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), Knowing, learning and instruction: Essays in honor of Robert Glaser. (pp. 393-451). Hillsdale, NJ: Erlbaum.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18, 32-42.

Butterfield, E.C., Albertson, L.R., & Johnston, J.C. (1993). On making cognitive theory more general and developmentally pertinent. Research on Memory Development: State-Of-The-Art and Future Directions: Conference proceedings 1993, Castle Ringberg, Germany, June 1993.

Clarke, V. A. (1990). Sex differences in computing participation: Concerns, extent, reasons and strategies. Australian Journal of Education, 34(1), 52-66.

Cox, D. A. & Berger, C. F. (1985). The Importance of Group Size in the Use of Problem-Solving Skills on a Microcomputer. Journal of Educational Computing Research, 1(4), 459-68.

Cunningham, D. J. (1992). In Defense of Extremism. Educational Technology, (31)9, 26-27.

Dalton, D. W., Hannafin, M. J., & Hooper, S. (1989). Effects of individual and cooperative computer-assisted instruction on student performance and attitudes. Educational Technology Research and Development, 37(2), 15-24.

Dalton, David. (1990). The effects of cooperative learning strategies on achievement and attitudes during interactive video. Journal of Computer-Based Instruction, 17(1) 8-16.

Davis, J.K., & Cochran, K.F. (1989). An information processing view of field dependence. Early Childhood Development and Care, 51, 31-47.

Dede, Christopher. (1993). Evolving from multimedia to virtual reality. Educational multimedia and hypermedia, 1994: Proceedings of ED-MEDIA 93-World Conference on Educational Multimedia and Hypermedia. Association for the Advancement of Computing in Education. (pp. 123-130).

Duffy, T.M. & Jonassen, D.H. (1992). Constructivism and the technology of instruction: A conversation. Hillsdale, NJ: Lawrence Erlbaum.

Fosnot, C.T. (1992). Constructing constructivism. In Duffy & Jonassen (Eds.), Constructivism and the technology of instruction: A conversation. Hillsdale, NJ: Lawrence Erlbaum. pp. 167-176.

Frank, B.M., (1983). Flexibility of information processing and the memory of field-independent individuals. Journal of Research in Personality, 17, 89-96.

Frank, B.M., Keene, D. (1993). The effect of learners field independence, cognitive strategy instruction and inherent word-list organization on free-recall memory and strategy use. The Journal of Experimental Education, 62(1), 14-25.

Gagné, R.M. (1974). Essentials of Learning for Instruction. Hinsdale, IL: Dryden Press.

Gagné, R.M. (1987). Instructional technology foundations. Hillsdale, NJ: Lawrence Erlbaum.

Glaser, R. (1990). The reemergence of learning theory within instructional research. American Psychologist, 45 (1) 29-39.

Goodenough, D.R., (1976). The role of individual differences in field dependence as a factor in learning and memory. Psychological Bulletin, 83, 675-694.

Held, R.M. & Durlach, N.I. (1992) Telepresence. Presence: Teleoperators and Virtual Environments, 1(1), 109-112.

Sheridan, T.B. (1992) Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments, 1(1), 109-112.

Hoffman, H. G., Hullfish, K. C., & Houston, S. J. (in press). Virtual-Reality monitoring. In Proceedings of the 1995 IEEE Virtual Reality Annual International Symposium (VRAIS). IEEE.

Johnson, D., & Johnson, R. (1987). Learning together and alone. Englewood Cliffs, NJ: Prentice Hall.

Johnson, R. T., Johnson, D. W., & Stanne, M. B. (1985). Effects of cooperative, competitive, and individualistic goal structures on computer-assisted instruction. Journal of Educational Psychology, 77(6), 668-677.

Johnson, R. T., Johnson, D. W., & Stanne, M. B. (1986). Comparison of computer-assisted cooperative, competitive, and individualistic learning. American Educational Research Journal, 23(3), 382-392.

Kazdin, A.E. (1982). Single case research designs: Methods for clinical and applied settings. Oxford: Oxford University Press.

Kourisky, M.L. (1983). Mini-Society: Experiencing real-world economics in the elementary school classroom. Menlo Park, CA: Addison-Wessley.

Kumar, David. (1994). Hypermedia: a tool for alternative assessment? Educational Technology Training and Instruction, 31(1) 59-66.

Lachman, R., Lachman, J., & Butterfield, E.C. (1979). Cognitive psychology and information processing: An introduction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Linn, R., Baker, E., & Dunbar, S. (1991) Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

Loftin, B., Engelberg, M., & Benedetti, R. (1993). Applying virtual reality in education: A prototypical virtual physics laboratory". IEEE (0-8186-4910-0)

Merrill, D. (1983). Component Display Theory, Reigeluth, C.M. Instructional design theory and models. Erlbaum

Moshell, J.M., and Hughes, C.E. (1994, January). Shared Virtual Worlds for Education. Virtual Reality World, 2 (1), 63-74.

Moss, P. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research. 62(3), pp. 229-258.

Nelson, W.A., Harmon, S.W., Orey, M.A., Palumbo, D.B. (1993). Techniques for Analysis and evaluation of user interactions with hypermedia systems. In Ed-Media 1993: Proceedings of. (pp. 585-588)

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

Newell, A. & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Novak, J. D. & Gowin, D. B. (1984). Learning how to learn. Cambridge: Cambridge University Press.

Ogborn, J. (1994). The design of exploratory and expressive learning environments. In R. Lewis and R. Mendelsohn (Eds.), Lessons from learning. Proceedings of the International Federation for Information Processing (IFIP) Working Conference on Lessons From Learning (pp. 125-135).

Pipho, C. (1992, April). The impact of a national test at the state level. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Quellmalz, E. S. (1991). Developing criteria for performance assessments: The missing link. Applied Measurement in Education, 4(4), 319-331.

Reeves, T. C. (1986). Research and evaluation models for the study of interactive video. Journal of Computer-Based Instruction. 13(4), 102-106.

Reeves, T. C. (1993). Pseudo-science in computer-based instruction: The case of learner control research. Journal of Computer-Based Instruction. 20(2), 39-46.

Regian, J. W., & Shebilske, W.L. (1992). Virtual reality: an instructional medium for visual-spatial tasks. Journal of Communication, 42(4), 136-149.

Resnick, L. (1987). Learning in school and out. Educational Researcher, 16, 13-20.

Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. Review of Educational Research, 64(4) 479-530.

Royer, J. M; Cisero, C.A. & Carlo, M.S. (1993). Techniques and procedures for assessing cognitive skills. Review of Educational Research, 63(2) 201-243.

Shavelson, R., Baxter, G. and Pine, J. (1991). Performance assessment in science. Applied Measurement in Education. 4(4), 347-62.

Sherwood, R.D., Kinzer, C., Hasselbring, T., & Bransford, J. (1987). Macro contexts for learning: Initial findings and issues. Journal of Applied Cognition, 1, 93-108.

Slavin, R. E. (1986). Using student team learning (3rd ed.). Baltimore: Johns Hopkins University, Center for Research on Elementary and Middle Schools.

Stephenson, S. D. (1991). The Effect of Instructor-Student Interaction on Achievement in Computer-Based Training (CBT). Interim Technical Paper for Period April 1990-February 1991. Air Force Office of Scientific Research, Washington, D.C.

Sternberg, R. (in press). For whom does The Bell Curve toll? It tolls for you. The New Republic.

Suchman, L., & Trigg, R. (1991). Understanding practice: Video as a medium for reflection and design. In J. Greenbaum & M. Kyung (Eds.) Design at work: Cooperative design of computer systems. Hillsdale, NJ: Lawrence Erlbaum.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. American Educational Research Journal, 31 (2), pp. 231-262.

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Webb, N. M., Ender, P., & Lewis, S. (1986). Problem-solving strategies and group processes in small groups learning computer programming. American Educational Research Journal, 23, 243-261.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, May, 703-713.

Wiggins, G. (1992). Create tests worth taking. Educational Leadership, 49(8), 26-33.

Winn, W. & Bricken, W. (1992). Designing virtual worlds for use in mathematics education: The example of experiential algebra. Educational Technology, 32 (12) 12-19.

Winn, W. & Bricken, W. (1992). Designing virtual worlds for use in mathematics education: The example of experiential algebra. Educational Technology, 32 (12) 12-19.

Winn, W. (1987). Instructional Design and Intelligent Systems: Shifts in the Designer's Decision-Making Role. Instructional Science, 16(1), 59-77

Winn, W.D. (1992). The assumptions of constructivism and instructional design. In Duffy & Jonassen (Eds.), Constructivism and the technology of instruction: A conversation. Hillsdale, NJ: Lawrence Erlbaum. pp 177-182.

Winn, W. (1993) *A Conceptual Basis for Educational Applications of Virtual Reality*. (Human Interface Technology Laboratory Technical Report #R-93-9). Seattle, WA: Human Interface Technology Laboratory Winograd, T., & Flores, F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, NJ: Ablex Publishing Co.

Yager, S., Johnson, D. W., & Johnson, R. T. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. Journal of Educational Psychology, 77(1), 60-66.